

# Stationary Component in Stock Prices: A Reappraisal of Empirical Findings

Haitham A. Al-Zoubi

*United Arab Emirates University, UAE*

Aktham Maghyereh

*United Arab Emirates University, UAE*

This paper re-examines the issue of mean reversion in stock prices by incorporating the structural break effect in the long horizon regression. Before adjusting for structural break, the paper finds that previous studies understate the evidence of mean-reversion. The understatement is mainly due to the clustering heteroskedasticity and autocorrelation in the overlapping returns. After adjusting for structural break(s), no evidence of predictability for value-weighted returns has been documented. However, stronger evidence of mean reversion in stock prices is documented for equally-weighted portfolios. The reverse effect of structural break can be explained by the switch to mean aversion in the last subperiod of value-weighted portfolios while no such switch in equally weighted portfolios (JEL: G1, C22).

**Keywords:** moving blocks bootstrap, mean reversion, structural change, long-horizon regressions.

## I. Introduction

The long-horizon predictability conditional on past returns has been the focus of many studies of market efficiency. Based on the variance ratio test of Cochrane (1988), Poterba and Summers (1988) and Lo and MacKinlay (1988) find that the variance of returns increases at a rate, which is less than proportional to the holding period. This implies that a substantial part of the variance in monthly returns is due to a predictable component. Fama and French (1988), using univariate autoregression of multiyear returns, also present evidence of mean reverting behavior and conclude that about 35% of stock variation is

---

\* We would like to thank Neal Maroney and Gerald Whitney for their guidance and helpful comments. All remaining errors are ours.

predictable from past returns. Further, Kim, Nelson, and Startz (1991), use stratified randomization to account for non-normality in the data, and find that mean reversion is only a pre-war phenomenon. The autoregression parameter flips to be positive for the post-war period (mean aversion). Finally, McQueen (1992) explicitly solves the problem of heteroskedasticity resulting from the commutative overlapping returns using a randomized Generalized Least Squares test. According to his test, he cannot reject the hypothesis of random walk, and finds that mean reversion is only statistically significant at the 3- and 4-year horizons in the 1926 to 1946 subperiod.

This paper re-examines the issue of mean reversion by incorporating the effect of a structural break on the parameter vector. A formal test of structural break was not taken into account when the previous tests were conducted. This extension is critical since the shift in the parameter estimate may be large enough to affect the value of parameter estimates and the standard error of the regression, which may give wrong inference about the random walk hypothesis. Another extension is the use of the Moving Blocks Bootstrap, hereafter *MBB*, of Fitzenberger (1998) in solving for the bias of the autoregression parameter and correcting for both the heteroskedasticity and autocorrelation problems. Fitzenberger (1998) established the ability of the *MBB* to provide asymptotic refinements when the errors are both heteroskedastic and autocorrelated. It is shown that the *MBB* heteroskedastic and autocorrelated consistent covariance matrix, hereafter *HAC*, is equivalent to the Barlett kernel suggested by Newey and West (1987) and performs better when the samples exhibit temporal dependence.<sup>1</sup>

The remainder of the paper is as follows. Section II reviews previous studies and shed lights on the complications of testing mean reversion. Section III provides the methodology and model estimation. Section IV outlines the data used. Section V tests the hypothesis of mean reversion and explains the results. Section VI concludes.

## II. Previous Studies and Structural Break in Stock Return

As has been noted by Summers (1986), there has been a lack of power in testing the random walk hypothesis using traditional tests. Fama and

---

1. Anderws (1991) shows that Barlet kernel *HAC* matrix of Newey and West behaves poorly when the data exhibits temporal dependence. This problem is more critical when the least square is used.

French (1988), hereafter *F&F*, find evidence of mean reversion in stock returns over long horizons (i.e., in excess of 18 months). They estimate an autoregression where the return over the interval  $t - T$  to  $t$ , call this  $R_{t-K,t}$ , is correlated with  $R_{t,t+K}$ :

$$R_{t,t+K} = \alpha + \beta_K R_{t-K,t} + \varepsilon_t \quad (1)$$

They consider return horizon  $K$  from one to ten years. They find little or no autocorrelation, except for holding periods of between  $K = 2$  and  $K = 7$  years for which  $B_T$  is less than zero. There was a peak at  $K = 5$  years when  $B = -0.5$ , indicating that a 10 percent negative return over five years is, on average, followed by a 5 percent positive return over the next five years. The  $R^2$  in the regression for the three to five-year horizons are about 0.35. Therefore, the mean reversion ( $B < 0$ ) is consistent with that from the anomalies literature where a “buy low, sell high” strategy earns persistent positive profits.

Four issues relating to autoregression test are now in order. First, the small sample bias of the *OLS* estimates of  $B_T$  since the true mean of the predictor is unknown in the finite sample. Kendall (1954) showed that the bias in the *OLS* estimate of  $BK$  is decreasing with the sample size and increasing with the value of the point estimate, in particular:

$$E(\hat{\beta}_K - \beta_K) = -(1 + 3\beta_K)/n + O(n^{-2}) \quad (2)$$

*F&F* correct for the bias in the least square estimate using Monte Carlo simulations, their results show that the bias is not large enough to affect the results. Using randomization simulation of Noreen (1989), Kim, Nelson, and Startz (1991) get the same results since the distribution does not affect the parameter estimate.<sup>2</sup>

Another issue concerns the methods used to correct for heteroskedasticity and autocorrelation in overlapping multiperiod returns. The problem of autocorrelation and heteroskedasticity leads to an inefficient least square estimate since the standard error is bias and inconsistent.<sup>3</sup> *F&F* use the method of Hansen and Hodrick (1980),

---

2. One of the differences between Monte Carlo simulation and randomization simulation is that the former assumes normality in random variables. The advantage of randomization over Monte Carlo is clear in the hypothesis testing if the return is not normally distributed.

3. Standard error of the regression is downward bias if errors are correlated and upward bias if they are heteroskedastic.

hereafter *HH*, which adjusts the autocorrelated standard error with  $MA(K - 1)$  error structure where  $K$  is the return horizon. In the case of multiperiod returns this method may be inappropriate since positive definiteness of the covariance matrix may not exist. Also they try to adopt White (1980) heteroskedastic consistency variance estimator with *HH* to solve for both problems. They find that the  $t$ -ratios are more dispersed than those of *HH*. Thus, they report the  $t$ 's based on *HH* alone. Richardson and Stock (1989) assume that the stock returns variance is stationary. Based on this assumption, they adjust the standard error for a form of stationary conditional heteroskedasticity. The resulting standard errors are so large as to provide a test statistic close to zero regardless of the point estimates.<sup>4</sup> However, Turner, Startz, and Nelson (1989) provide evidence that the variance of the long-horizon return is nonstationary. Moreover, McQueen (1992) explicitly solves for the clustering heteroskedasticity results from overlapping returns using the *GLS* randomization test. According to his results, he cannot reject the random walk hypothesis. However, there are two problems with McQueen's method used to solve the problem of heteroskedasticity in the overlapping returns.

First, the historical *GLS* results are inefficient since the first observation in the estimation is dropped and the problem will be more severe as the return horizon becomes longer. For example, in estimating the 10-year horizon returns he drops the first 120 observations from the monthly returns.

Second, the use of the Depression/World War II period as a weight in the *GLS* randomization test is adopted without a formal test of structural change in the returns variance, which may overstate the result of no mean reversion.

The third dimension of the problem is the existence of a structural break in the stock return horizon, Fama and French argue "the evidence is clouded by statistical issues (changing parameters, heteroskedasticity) that such a long time raises". Following Perron (1989), it is known that apparent persistence in the data could be the result of unmodeled structural breaks in the underlying data process. Therefore, the test for the existence of the stationary component in stock returns may be weakened in favor of accepting the random walk since the transitory component is no longer stationary, from  $t$  to  $t + k$ , even it is stationary about the break.

---

4. Richardson (1993) argues that properly adjusting for finite sample bias, heteroskedasticity, and autocorrelation in univariate autoregression may reverse many of the inferences of *F&F* (1988).

### III. Methodology

This section describes the moving sums standardized residuals (*MOSUM*) test introduced by Chu, Hornik, and Kuan (1995a) and the moving estimate (*ME*) test introduced by Chu, Hornik, and Kuan (1995b) to test for endogenous structural break(s) and the moving blocks bootstrap (*MBB*) of Fitzenberger (1998) adopted in the paper for testing the long-horizon predictability in stock returns. First, this section briefly reviews structural break tests. Next, the *MBB* is introduced.

#### A. Structural Change Test

Recent advances in the econometrics of structural break now allow for robust tests of parameter instability without assuming exogenous change points. Among the proposed structural break tests, the moving sums standardized residuals (*MOSUM*) test introduced by Chu, Hornik, and Kuan (1995a) and the moving estimate (*ME*) test introduced by Chu, Hornik, and Kuan (1995b) are particularly attractive since they dominate the optimal *average F*- and *exponential F*-tests of Andrews and Ploberger (1994) under various alternatives of double structural change.<sup>5</sup> Rather, the *MOSUM* identification of the structural break is determined endogenously based on the moving sums of fixed number of residuals by the probability of a Wiener process crossing a pair of linear boundaries.<sup>6</sup> Hence, the *MOSUM* test becomes more sensitive to temporal parameter instability than the *CUSUM* test based on *OLS* residuals proposed by Ploberger and Kramer (1992) since the later incorporates more residuals.<sup>7</sup> This property of the testing procedure is particularly attractive in the long-horizon regression where the overlapping returns carry out the points of structural breaks across the horizon.

Unlike *CUSUM*, the *MOSUM* considers testing procedures where

---

5. One of the disadvantages of the *Ave* (Average) and *Exp* (Exponential) tests of Andrews and Ploberger (1994) is that they perform quite poorly when an incorrect alternative is assumed.

6. Viceira (1997) uses the recursive and reverse recursive tests based on the spirit of Ploberger, Kramer, and Kontrus (1989), Brownian motion process to test for parameter stability for the long horizon regression. He argues that the fluctuation tests based on the Brownian motion process is more appropriate than the sequential tests in the case of testing for mean reversion.

7. Moreover, Chu, Hornik, and Kuan (1995a) show that *MOSUM* test outperforms Ploberger and Kramer (1992) test in terms of empirical power.

the empirical fluctuation process contains the sum of residuals in a data window whose size is determined by a bandwidth parameter  $h \in (0,1)$  and which is moved over the whole sample period. Chu, Hornik, and Kuan (1995a) suggest considering the following *OLS* based *MOSUM* empirical fluctuation process:

$$M_n^0(t/h) = \frac{1}{\hat{\sigma}\sqrt{n}} \left( \sum_{i=\lfloor N_n t \rfloor + \lfloor nh \rfloor}^{\lfloor N_n t \rfloor + \lfloor nh \rfloor} \varepsilon_i \right) \quad (0 \leq t \leq 1-h) \quad (3a)$$

$$= W_n^0 \left( \frac{\lfloor N_n t \rfloor + \lfloor n, h \rfloor}{n} \right) - W_n^0 \left( \frac{\lfloor N_n t \rfloor}{n} \right), \quad (3b)$$

where  $N_n = (n - \lfloor nh \rfloor)/(1 - h)$ . As appears in (3b), the limiting process of the *MOSUM* empirical fluctuation is a Brownian bridge.

Instead of defining the fluctuation process on the basis of the residuals, the *ME* test fluctuation process is based on estimates of the regression coefficients and determined by the increments of a vector of Brownian bridge. This gives the *ME* process of the following  $K$ -dimensional Brownian bridge specification:

$$Z_n(t/h) = \left( \frac{\sqrt{\lfloor nh \rfloor}}{\hat{\sigma}\sqrt{n}} \right) \left( X^{(\lfloor nr \rfloor, \lfloor nh \rfloor)T} X^{(\lfloor nr \rfloor, \lfloor nh \rfloor)T} - \hat{B}^{(n)} \right), \quad (4)$$

where  $0 \leq t \leq 1-h$  and  $X$  is the vector of regressors.

Chu, Hornik, and Kuan (1995b) show that the *ME* test is consistent, has nontrivial local power, and outperforms the fluctuation test of Ploberger, Kramer and Kontrus (1989) and the sequential tests of Andrews (1993) and Andrews and Ploberger (1994) when parameters display temporal instability. Moreover, the test is robust to heteroskedasticity and autocorrelation. More recently, Chu, Stinchcombe, and White (1996) and Leisch, Hornik, and Kuan (2000) develop a test for monitoring structural breaks when there is a new arrival of observations to see if a change in regression coefficients occurs after the new arrival. This could be attractive to insure that the conclusion of this study is not derived by a different sample period since our sample period (1926 – 2001) incorporates more data than others.<sup>8</sup>

---

8. The sample period of Fama and French (1988) is from January 1926 to December

For the empirical analysis, this paper constructs three tests; the first based on *OLS* residuals (*MOSUM*), the second based on the parameters estimate (*ME*), and the last one is a forward looking test based on monitoring the *ME* test.<sup>9</sup> For the monitoring test, the paper chooses the sample period of *F&F* from January 1926 to December 1985 as a historical period, and the sample period from January 1986 to December 2001 as newly arrived data.

### B. Moving Blocks Bootstrap (*MBB*)

Efron (1979) established the theory of Bootstrap to reduce the finite-sample bias and yield an approximation to the distribution of an estimator or test statistic that is at least as accurate as the approximation obtained from first-order asymptotic theory. Efron suggests two basic approaches to bootstrap the linear regression. One is to first fit the model and apply the bootstrap to the residuals. The resulting covariance estimate for this case is similar to that of the point estimate, except for the degrees of freedom adjustment (see, Efron [1982]), for further discussion). A second approach, with more general applicability, is the Design-Matrix-Bootstrap, hereafter *DMB*. Under this approach, the entire vector of the dependent variable and the regressors are bootstrapped. So *DMB* provides a heteroskedastic consistent covariance matrix, hereafter *HC*, since the errors are not resampled.<sup>10</sup> Singh (1981) contradicts the asymptotic validity of the bootstrap covariance estimator in the case of autocorrelation. He notes that both approaches cannot provide an autocorrelated consistent estimate, hereafter *AC*, in the linear regression case. Wu (1986) claims that the Jackknife approach performs better than the bootstrap particularly in the case of heteroskedastic residuals. Lui (1988) introduces the wild bootstrap to solve the bias of the autoregression parameter. Mammen (1993) establishes the ability of the wild bootstrap to provide asymptotic refinements when the errors

---

1985. McQueen (1992) expands the sample of Fama and French (1988) backward by incorporating the monthly Cowles commission (1939) data and forward by incorporating the monthly data of the years 1986 and 1987. The comparison of our results is based on the data after January 1926.

9. For the monitoring test this paper follows the methodology suggested by Leisch, Hornik, and Kuan (2000).

10. Efron and Tibshirani (1986) claim that both approaches are asymptotically equivalent presumably when the covariance is assumed to be chosen from a probability distribution.

are heteroskedastic. Using Monte Carlo experimentation, Hall, Horowitz, and Jing (1995) show that wild bootstrap performs better than any version of *HC* estimator.<sup>11</sup>

The *MBB* approach adopted in this study is introduced independently by the work of Kusch (1989) and Liu and Singh (1992) to provide a *HAC* standard errors equivalent to the *NW* estimator. The *MBB* can be implemented by dividing the data into blocks and the bootstrap sample is obtained by sampling blocks randomly with replacement. To describe the method of blocking the data, let the sample consist of observations  $\{y_i, x_i : i = 1, \dots, T\}$ . With overlapping blocks of size  $b$  and length  $L = i + b - 1$ , block  $B_i$  is the entire vector of blocks  $(B_i^y, B_i^x)$  with  $B_i^y = (y_i, \dots, y_{i+b-1})$ ,  $b \times 1$  vector, and  $B_i^x$ , a  $b \times k$  matrix of regressors. The bootstrap resample  $\{(y_i^*, x_i^*), \dots, (y_1^*, x_1^*)\}$  of size  $l = bm$  is generated by drawing  $m$  *iid* ( $b \times [k + 1]$ ) blocks, call them  $\{Z_1, \dots, Z_m\}$ , from  $q = T - b + 1$  blocks  $B_i$ . The *MBB* resample is formed by laying  $\{Z_1, \dots, Z_m\}$  end-to-end in the order sample. The resulting vector,  $\{Z_{\tau v} = (y_i^*, x_i^*)\}$ , is the *MBB* sample, where  $\tau = [(i - 1)/b] + 1$  and  $v = i - b\tau$ .

Fitzenberger (1998) compares the performance of *MBB* with Barlett kernel suggested by Newey and West (1987) and quadratic spectral kernel (*QS*) suggested by Andrews (1991).<sup>12</sup> The Monte Carlo simulation shows that the *MBB* dominates both estimators in terms of variance bias and coverage properties especially when the data exhibit heteroskedasticity with increasing dependence i.e., the autocorrelation is rising.<sup>13</sup> The results show that the improvement in variance bias and undercoverage is monotonically increasing with block size. However, the improvement of the root mean square error (*RMSE*) is not monotonic with increasing  $b$ . Consistent with Hall, Horowitz, and Jing (1995), the results show that the block size has to grow at the rate  $T^{1/3}$  to minimize the *RMSE* of the variance estimate. This is identical to Andrews' (1991) optimal bandwidth.

---

11. Hall et al. (1995) compare the wild bootstrap with the jackknife approximation of the heteroskedasticity-consistent covariance matrix estimator of MacKinnon and White (1985) and the bootstrap method of resampling the dependent and the independent variables. The results show that the wild bootstrap provides a *t*-statistic with the lowest distortions.

12. Andrews (1991) discusses different Kernel *HAC* estimators used in the literature. He shows that quadratic spectral kernel covariance estimator (*QS*) dominates Barlett kernel suggested by Newey and West (1987) in terms of both *RMSE* and the true confidence interval performance.

13. Fitzenberger (1998) shows that the result is hold whether the *MBB* is centering around sample estimate or resample mean.

### C. Model Estimation

We construct three subperiods for each structural break in the non-overlapping returns for each horizon of the overlapping returns. First, the overlapping generating process does not have a break point. Second, the break point overlaps in the return horizon. Third, the overlapping returns become free of the break point.

The following *MBB* specification of the first order autoregression is used to test for the random walk hypothesis across subsamples:

$$R_{t,t+K} = \alpha_K + \sum_i^n C_i \alpha_{K,i} + \sum_{j=1}^m \beta_{K,D_j} D_j R_{t-k,t} + \varepsilon_t \quad (5)$$

where  $n$  and  $m$  is the number of structural breaks in the intercept and the slope of the long horizon regression, respectively.  $C = 1$  if  $t$  is in the  $i^{\text{th}}$  subhorizon and 0 otherwise,  $D = 1$  if  $t$  is in the  $j^{\text{th}}$  subhorizon and zero otherwise, and the subscripts  $\alpha_K$ ,  $\alpha_{K,i}$ ,  $\beta_{K,D_j}$  indicate the estimated parameters for each subhorizon  $i$  and  $j$ , respectively. The specification in (3) allows the slope to display different points of structural breaks than that of the intercept. This may be more convenient since we are interested in estimating the slope of the autoregression. The null hypothesis of interest is:

$$\sum_j^m \beta_{K,D_j} = 0$$

against the alternative of a break-mean reversion model.

## IV. The Data

The raw data for equally-and value-weighted returns, includes dividends, for all stocks traded on the New York Stock Exchange (NYSE) consist of end-of-month observations from January 1926 until December 2001 are collected from the Center for Research in Security Prices (CRSP). Following *F&F*, we transform the one-month returns into continuously compounded form. The resulting nominal returns are then adjusted for the inflation rate using the Consumer Price Index (CPI) to get the real returns. The overlapping monthly returns on each horizon are then calculated by summing up the real returns.

**TABLE 1. Moving Sums Standardized Residual (MOSUM) Test and Moving Estimate (ME) Test for the Hypothesis of no Structural Break in the First Order Autoregression**

	Return Horizon (K months)							
	12	24	36	48	60	72	96	120
Panel A: Equally-Weighted								
<i>MOSUM-OLS:</i>								
Empirical Fluctuation process	3.510 (0.021)	5.377 (0.013)	6.360 (0.014)	7.010 (0.009)	6.938 (0.009)	9.092 (0.007)	7.172 (0.009)	7.861 (0.008)
<i>ME:</i>								
Parameter Vector	3.753 (0.025)	5.454 (0.019)	6.485 (0.016)	7.651 (0.012)	7.844 (0.009)	5.992 (0.017)	7.835 (0.009)	8.291 (0.004)
Panel B: Value-Weighted								
<i>MOSUM-OLS:</i>								
Empirical Fluctuation process	2.650 (0.027)	4.335 (0.019)	5.928 (0.012)	7.197 (0.011)	7.584 (0.010)	7.462 (0.012)	7.942 (0.008)	8.401 (0.008)
<i>ME:</i>								
Parameter Vector	2.647 (0.020)	4.295 (0.021)	5.93 (0.017)	7.176 (0.015)	7.584 (0.014)	6.983 (0.015)	7.508 (0.014)	7.206 (0.015)

**Note:** Figures in parentheses are p-values.

## V. Empirical Results

In this section, the paper first reports the results of stability tests based on the empirical fluctuation test, individual parameters, and full parameter vector for the value- and equally-weighted portfolio real return. Next, the *MBB* estimates of the *F&F* autoregression for the overall sample period and different subperiods selected on the basis of the stability tests are reported.

### A. Model Stability Test

Table 1 reports the results of the structural break tests in the autoregression model for the value- and equally-weighted portfolio real returns. Following Chu, Hornik, and Kuan (1995a, 1995b), the paper sets the fixed number of residuals in a data window whose size is 20 percent of the effective sample when the *MOSUM* test is considered, and the duration of parameter change to 50 percent when *ME* test is considered.

Table 1 reports the results of structural break tests for the value-and equally-weighted overlapping returns for 1-, 2-, 3-, 4-, 5-, 6-, 8- and 10-year horizons sequentially. The findings in table 1 provide strong evidence that the autoregression test of the random walk hypothesis displays structural breaks for monthly and overlapping monthly returns over the effective sample period. The results are robust across the *MOSUM* and *ME* tests and all *p*-values indicate a rejection of the null hypothesis of parameter stability at 1% and 5% significance level. The results are analogous to the individual parameters (the constant and the parameter of lagged return) for the overlapping monthly returns. For one-month portfolio returns we find no evidence of intercept instability, non of the *p*-values below 5% significance level.

To gain some insight into the nature of the instability, figures 1-16 graph the values of *MOSUM* empirical fluctuation process supported by a pair of linear boundaries computed at 5% critical value for the null hypothesis of no structural change in the full parameter vector of overlapping monthly returns. For both the value-and equally-weighted overlapping returns, the maximum deviations in the statistic are occurred around the first quarter of 70's when the statistics sharply fall down across all horizons. Also, the autoregression parameter vector displays another structural break for the most of the horizons. For equally weighted returns, the second structural break in the 1-, 2- and

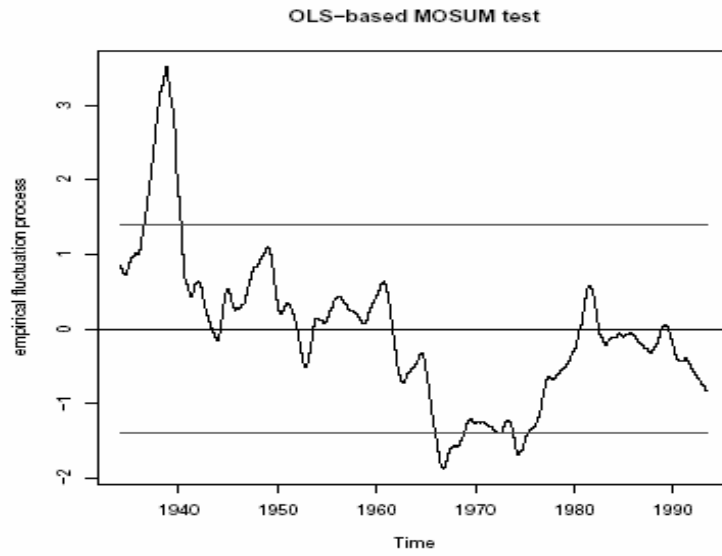


FIGURE 1—12-Month Equally Weighted Return Horizon

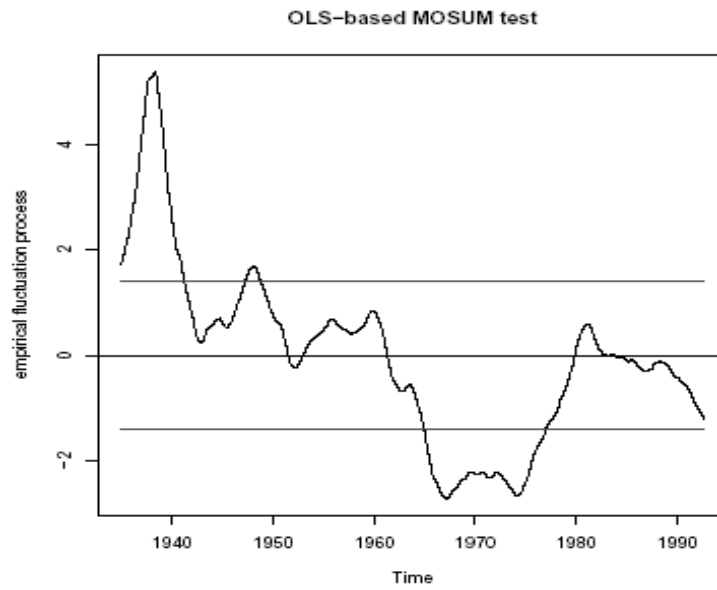


FIGURE 2—24-Month Equally Weighted Return Horizon

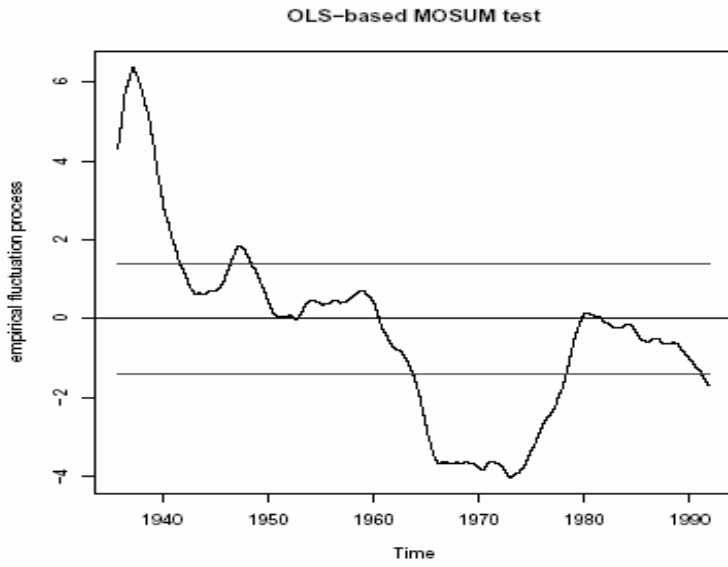


FIGURE 3—36-Month Equally Weighted Return Horizon

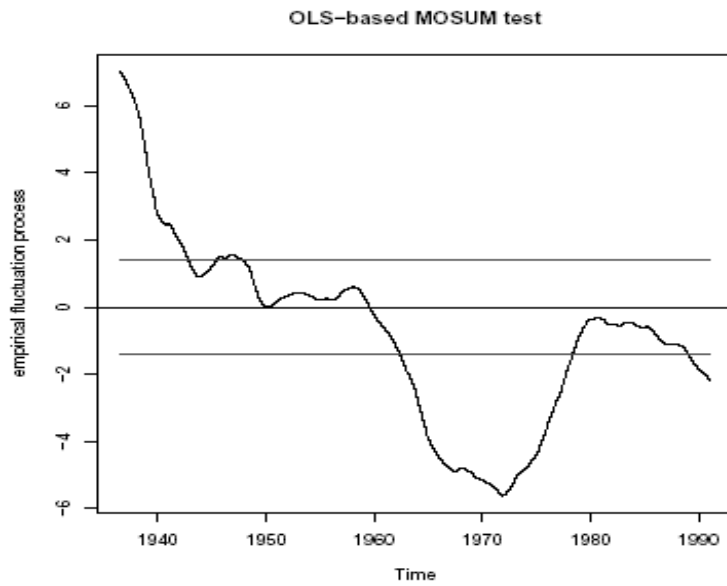


FIGURE 4—48-Month Equally Weighted Return Horizon

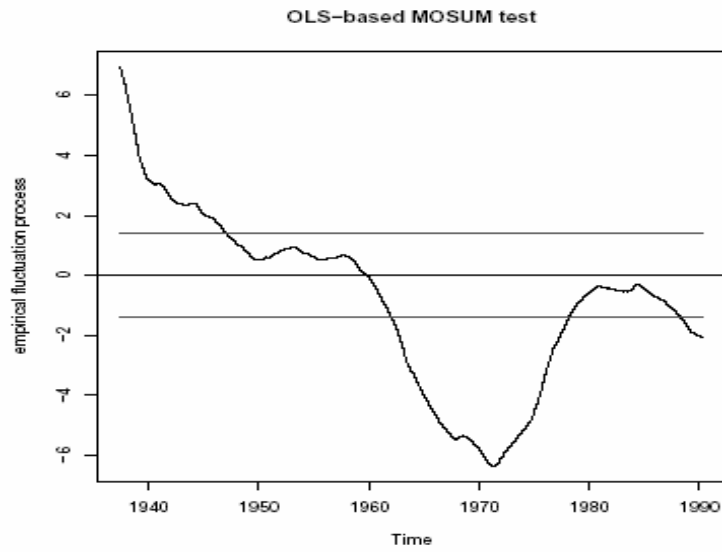


FIGURE 5—60-Month Equally Weighted Return Horizon

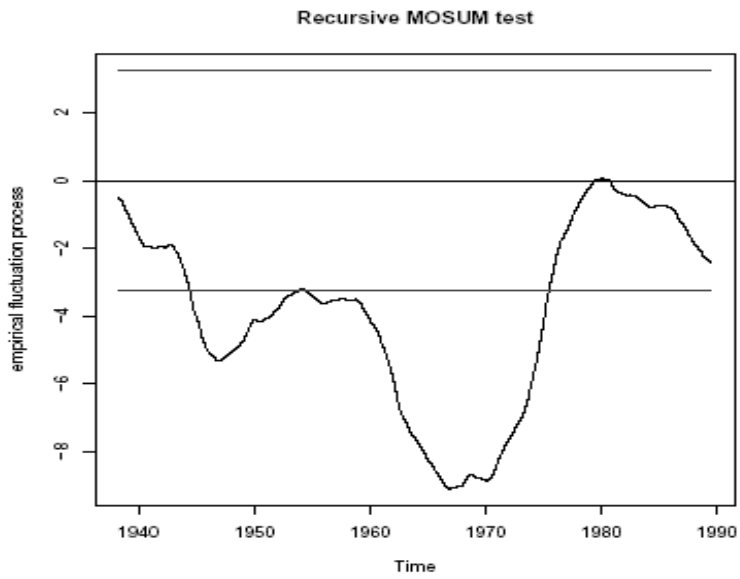


FIGURE 6—72-Month Equally Weighted Return Horizon

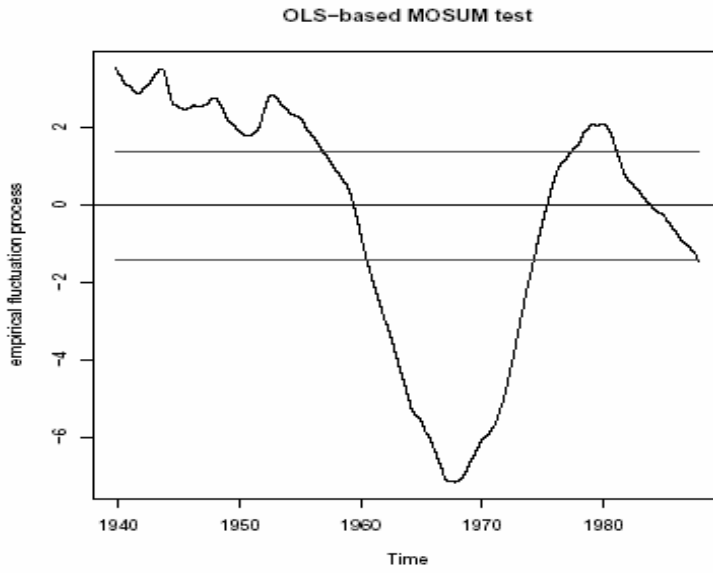


FIGURE 7—96-Month Equally Weighted Return Horizon

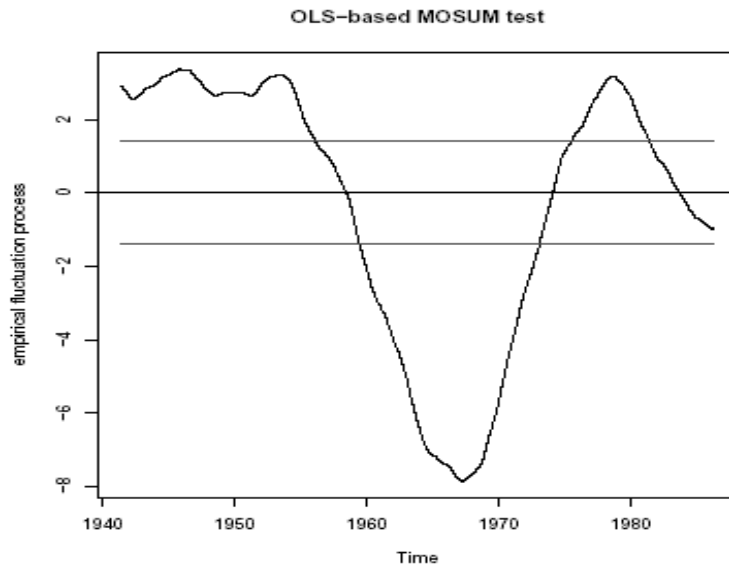


FIGURE 8—120-Month Equally Weighted Return Horizon

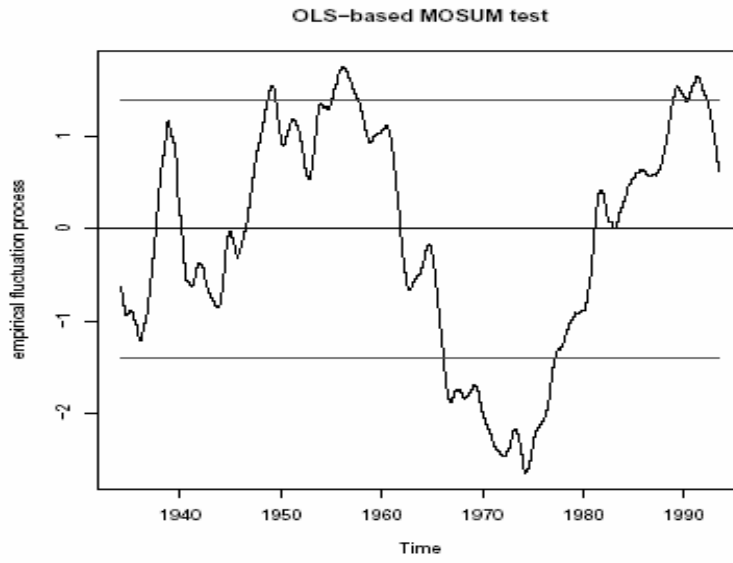


FIGURE 9—12-Month Value Weighted Return Horizon

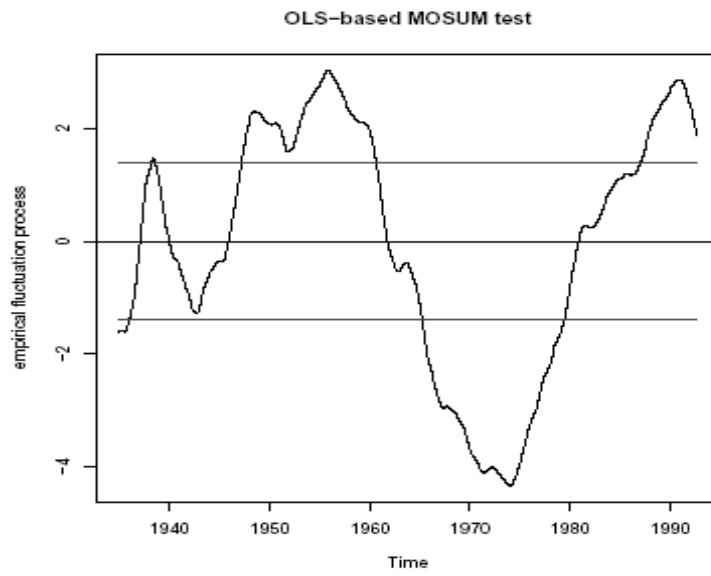


FIGURE 10—24-Month Value Weighted Return Horizon

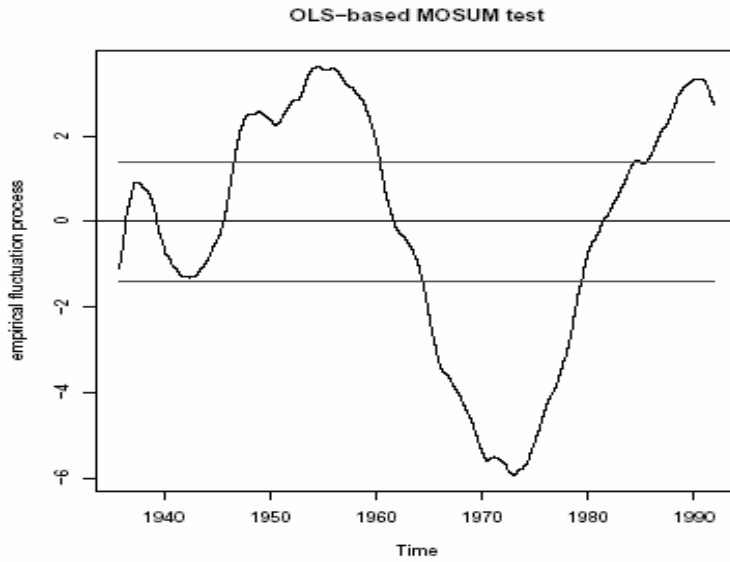


FIGURE 11—36-Month Value Weighted Return Horizon

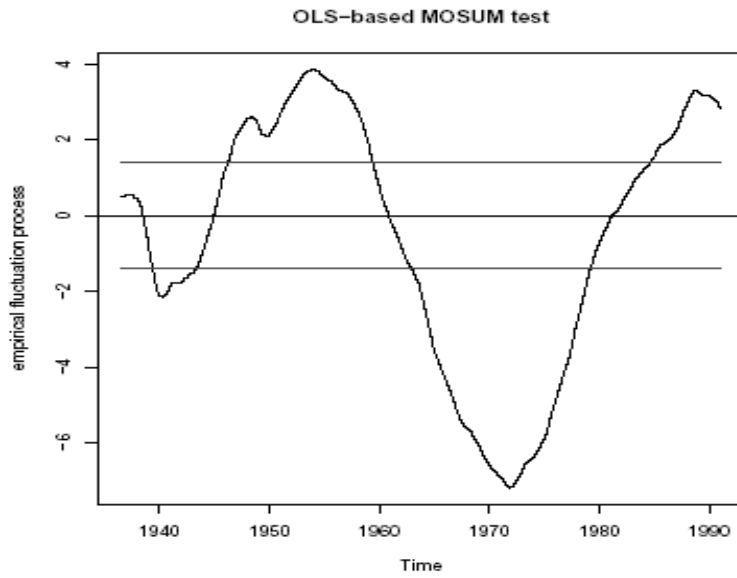


FIGURE 12—48-Month Value Weighted Return Horizon

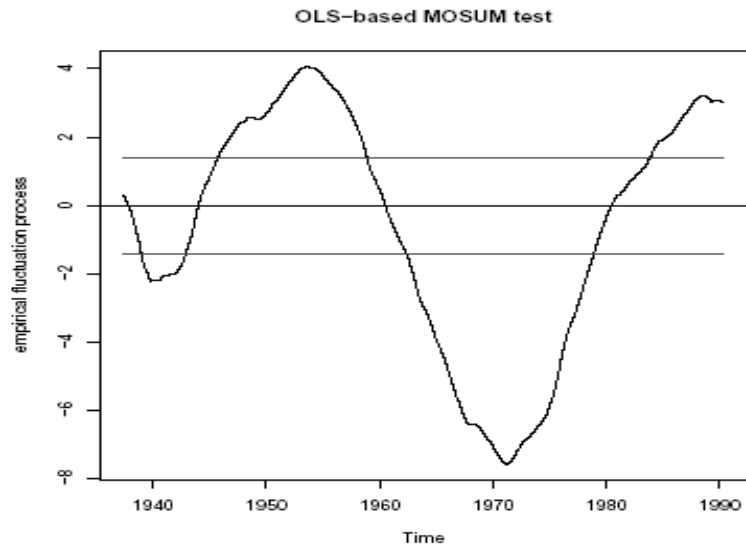


FIGURE 13— 60-Month Value Weighted Return Horizon

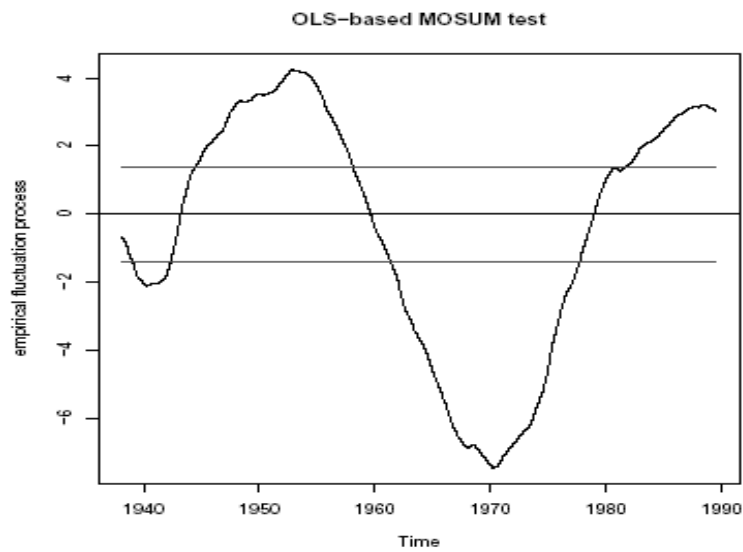


FIGURE 14— 72-Month Value Weighted Return Horizon

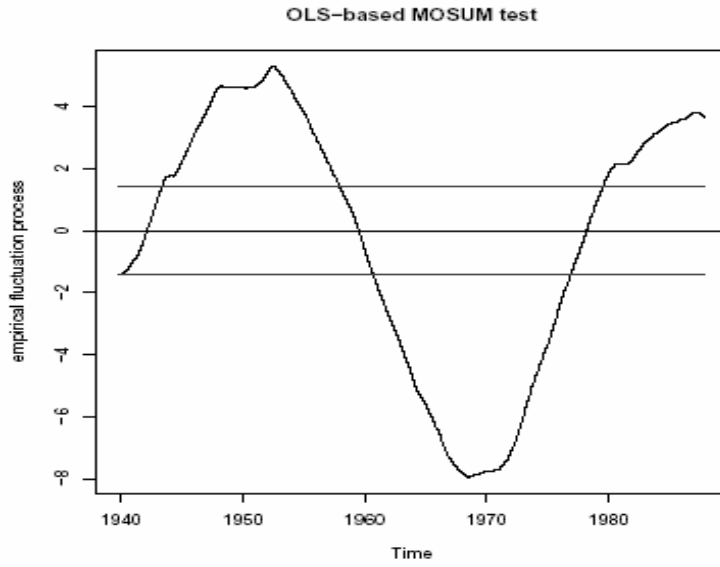


FIGURE 15—96-Month Value Weighted Return Horizon

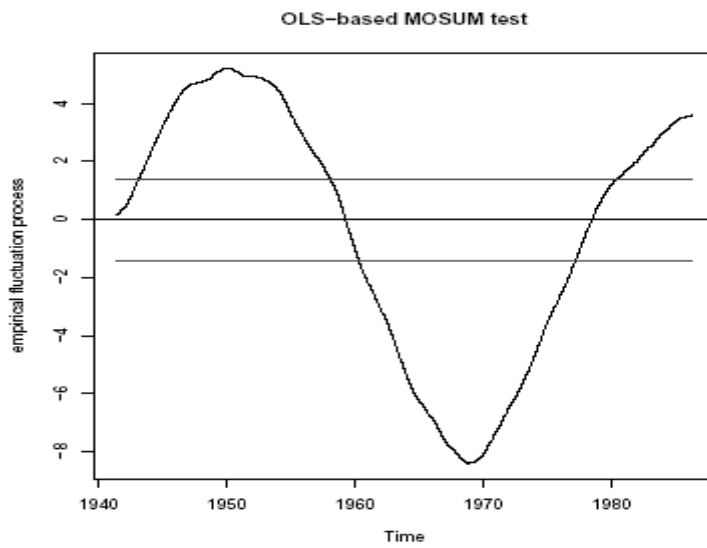


FIGURE 16—120-Month Value Weighted Return Horizon



3-year horizons occurred at the beginning of the 40's, the 4-, 5- and 6-year horizons displays only one structural break, and the second shift in 8- and 10-year horizons occurred around the end of 70's. For value weighted returns, the fluctuation process is clearer about the second structural break. The results show that the second shift is occurred around the beginning of 50's across all horizons.

Figures 17-32 graph the values of *ME* test for the individual parameters across the long horizon returns. The most critical finding in the graphs is that, in the most cases, the structural break deducted around the beginning of 70's is related to a shift in the intercept not to the slope of the autoregression. In other words, the *ME* test provides evidence that for the majority of the return horizons the date of structural shift in the slope of the model differs than that for the intercept. This result is very important when we test for random walk using the autoregression model since we are interested in the value of the slope rather than the intercept. For example, the value-and equally-weighted returns displays a structural break in the intercept across all horizon around the beginning of 70's but value weighted 1-, 3-, 4- and 5-year return horizons displays no structural break in the slope at that period. To summarize the results, table 2 reports the selected breakdates for the overlapping monthly returns on the basis of *MOSUM* and *ME* test statistics.

Three main results are revealed from the analysis. First, the main structural break dates for those series' occurred on the subhorizon where the depression period returns are not overlapped. The strongest peak and bottom for the considered tests indicate that the main shifts in entire parameter vector and individual parameters occur around the beginning of the 70's or around the mid 80's. Second, there is strong evidence that the overlapping returns display a second structural break where the depression period returns are overlapped. The *MOSUM* test points out that the shift occurred around the beginning of 50's and the *ME* test clarifies that shift in the 50's is generally related to the intercept while the shift in the slope of the autoregression occurs at the beginning of the effective sample (the mid 40's). Third, the structural breaks in the intercept are fixed, around the beginning of the 50's and 70's, across the horizons while the shift in the slopes are monotonic with horizon. Taken together, graphs 1-18 and tables 1 and 2 provide the date and the nature of the structural break for each single series.

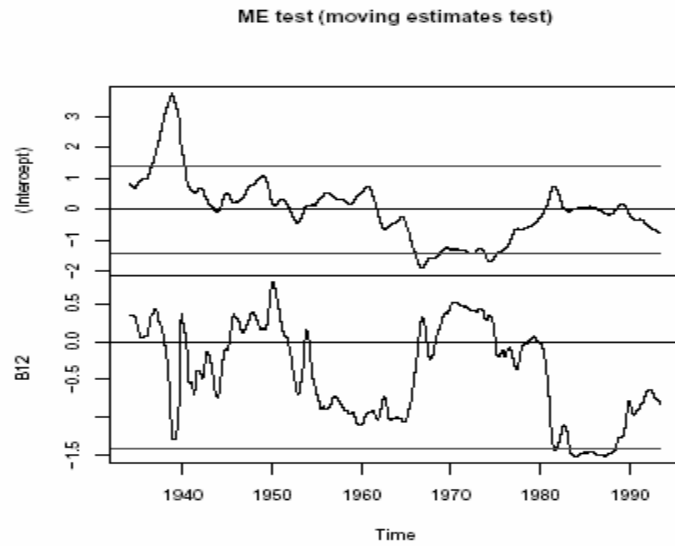


FIGURE 17—12-Month Equally Weighted Return Horizon

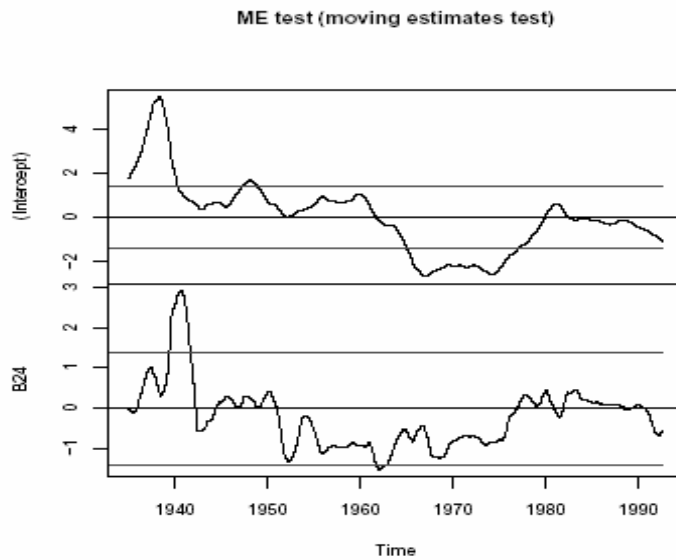


FIGURE 18—24-Month Equally Weighted Return Horizon

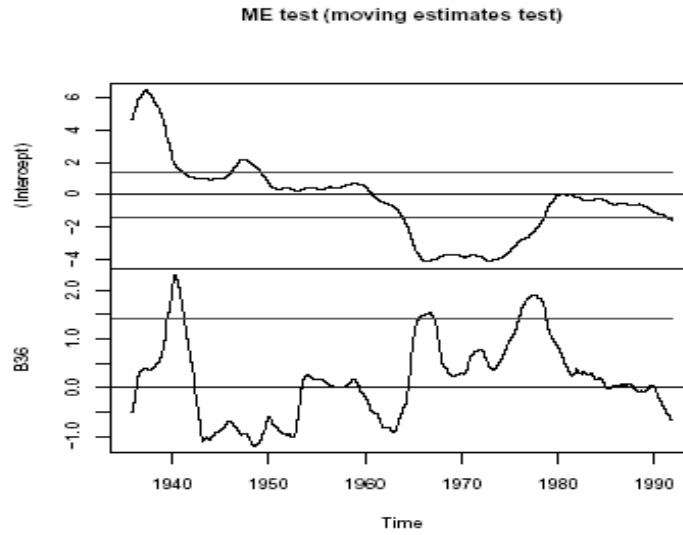


FIGURE 19—36-Month Equally Weighted Return Horizon

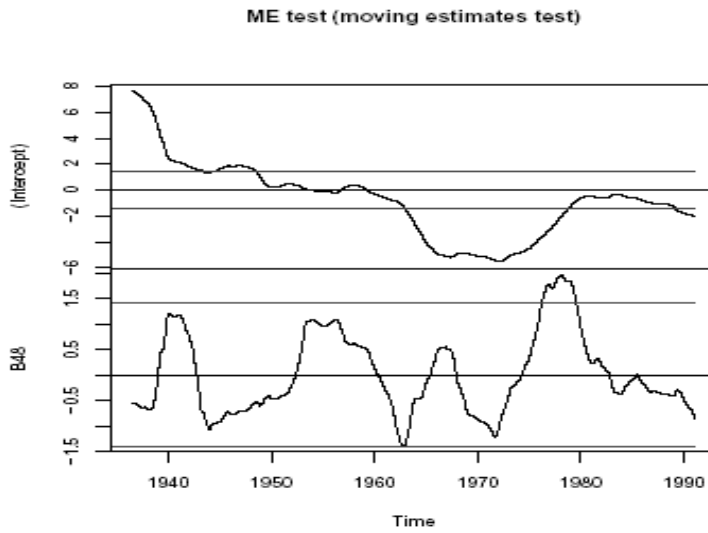


FIGURE 20—48-Month Equally Weighted Return Horizon

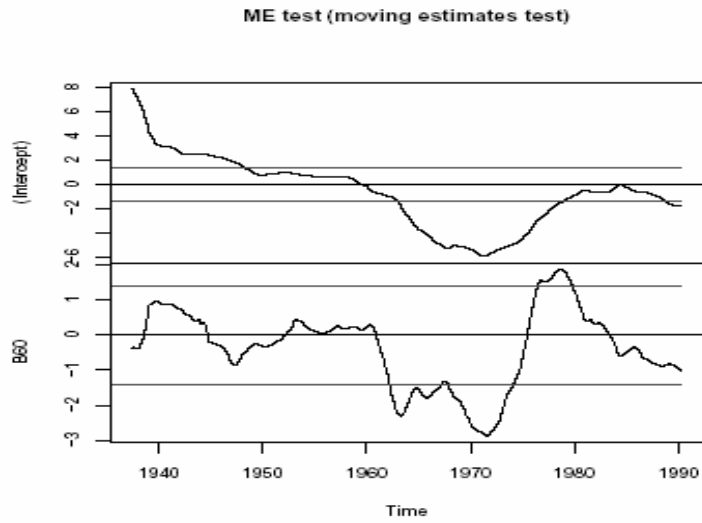


FIGURE 21—60-Month Equally Weighted Return Horizon

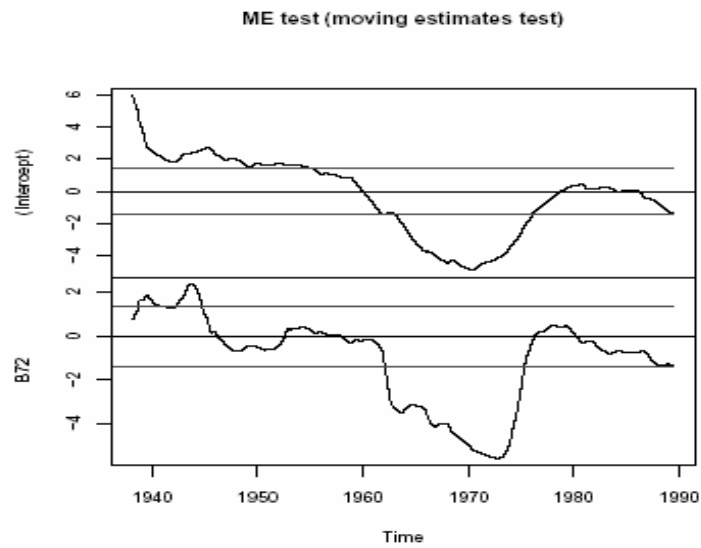


FIGURE 22—72-Month Equally Weighted Return Horizon

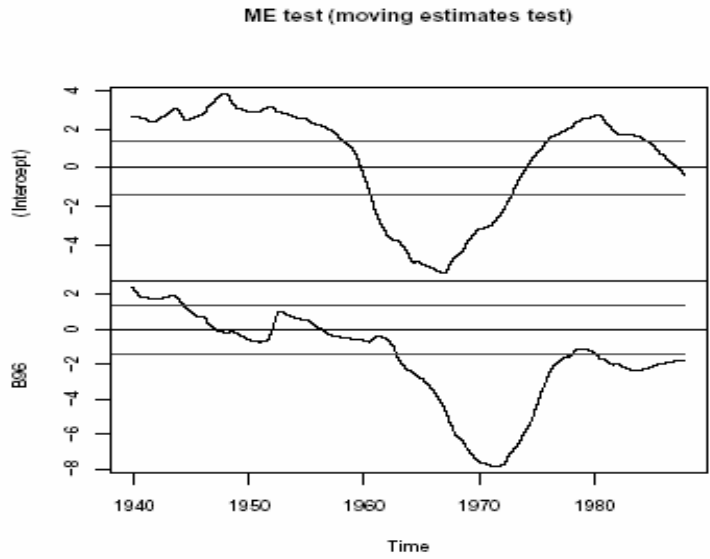


FIGURE 23—96-Month Equally Weighted Return Horizon

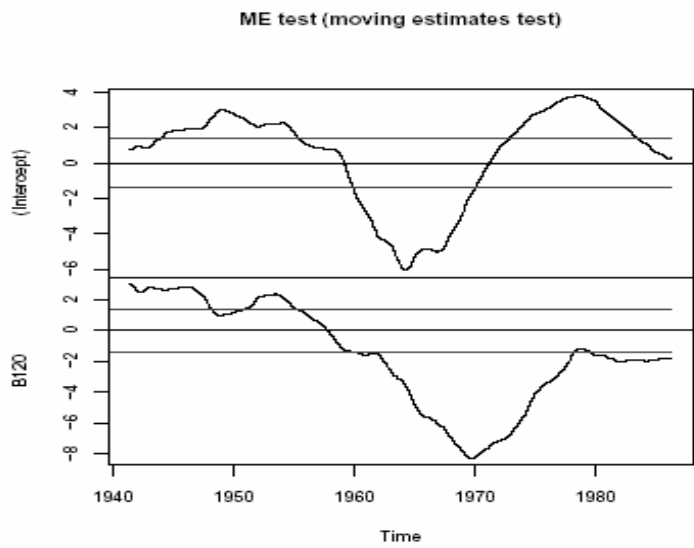


FIGURE 24—120-Month Equally Weighted Return Horizon

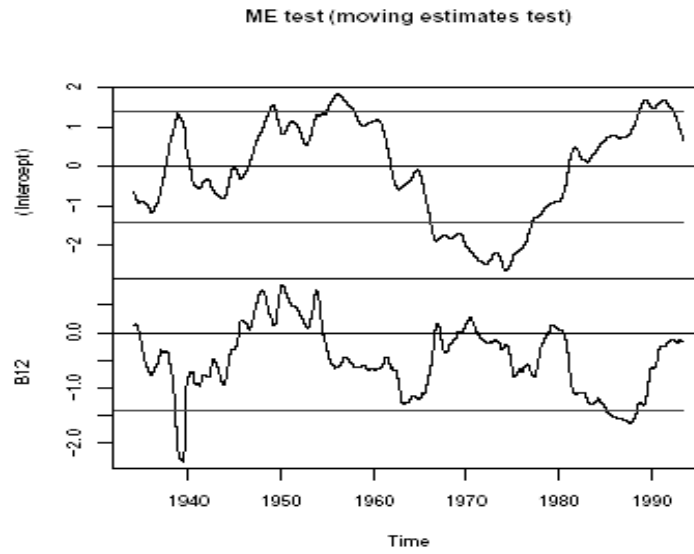


FIGURE 25—12-Month Value Weighted Return Horizon

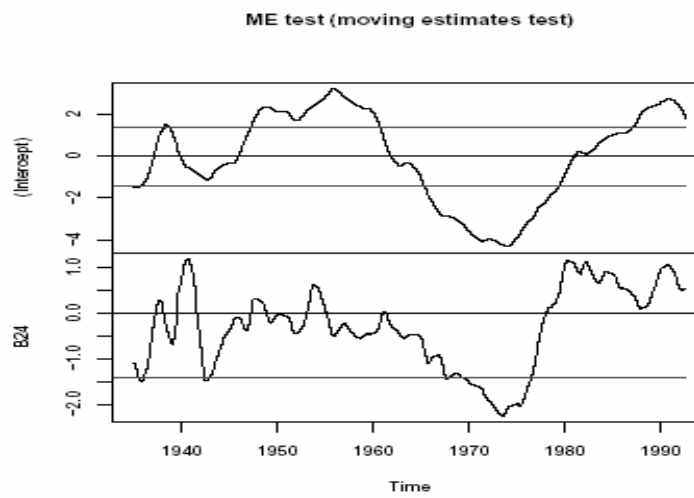


FIGURE 26—24-Month Value Weighted Return Horizon

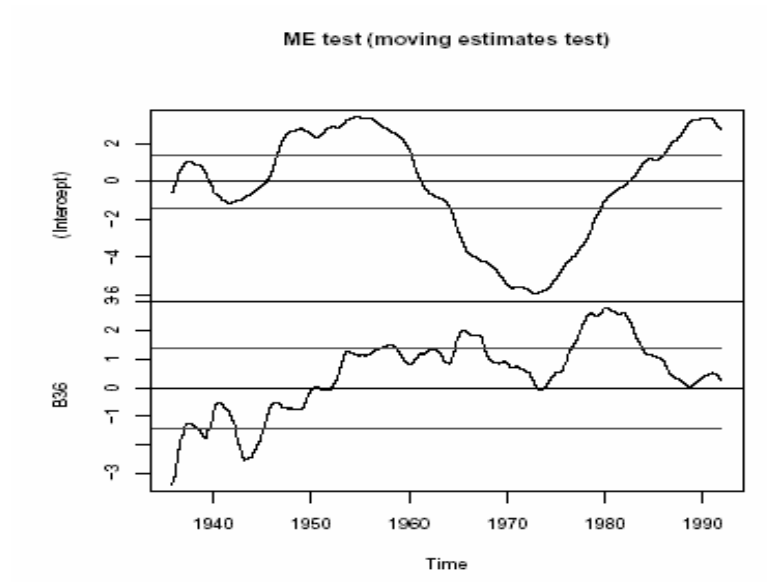


FIGURE 27—36-Month Value Weighted Return Horizon

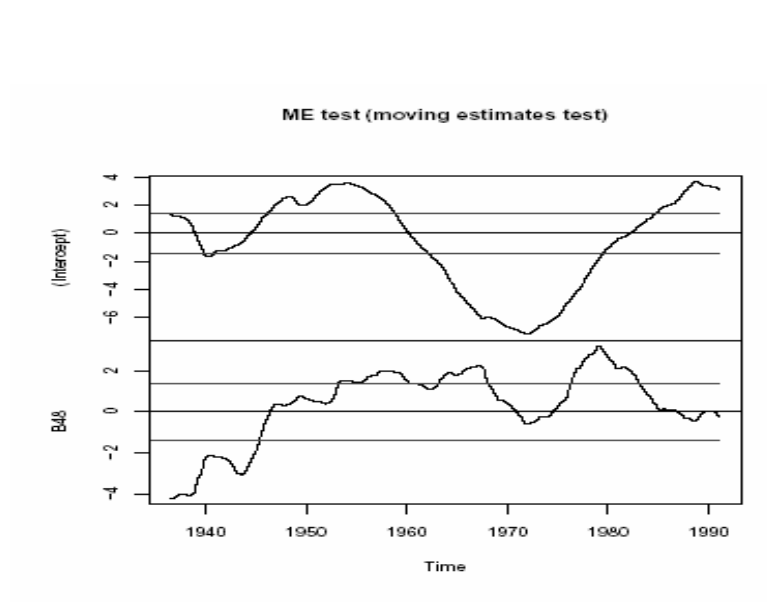


FIGURE 28—48-Month Value Weighted Return Horizon

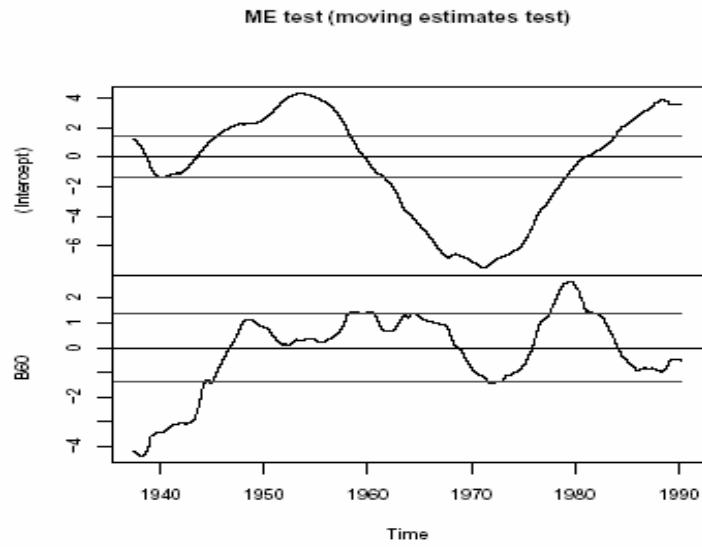


FIGURE 29—60-Month Value Weighted Return Horizon

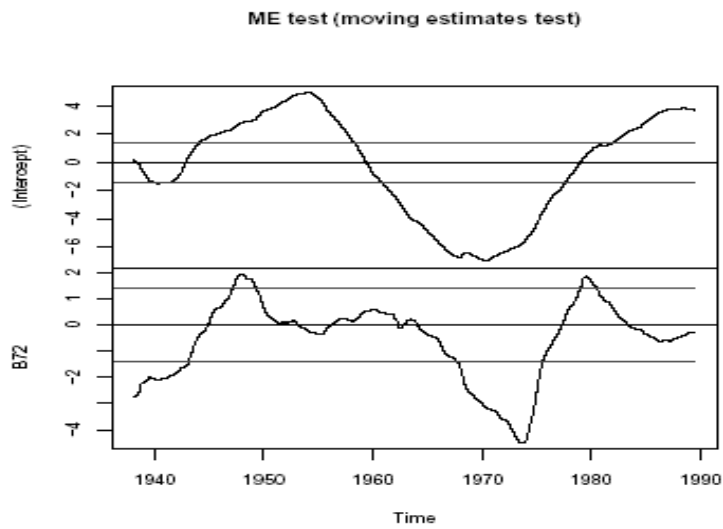


FIGURE 30—72-Month Value Weighted Return Horizon

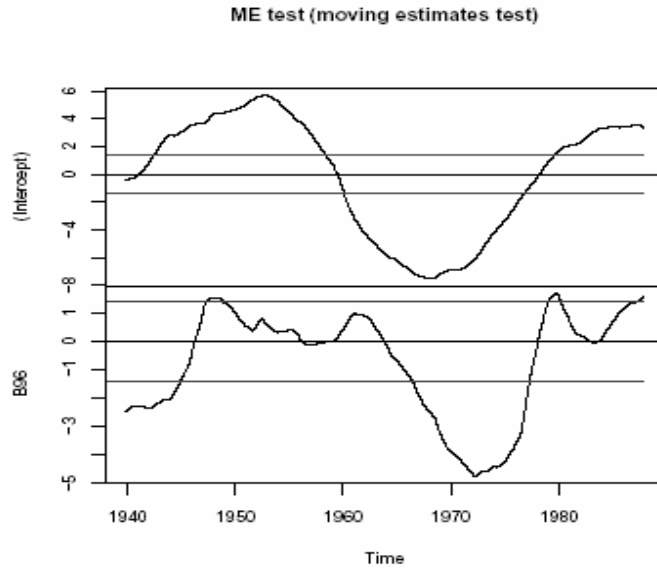


FIGURE 31—96-Month Value Weighted Return Horizon

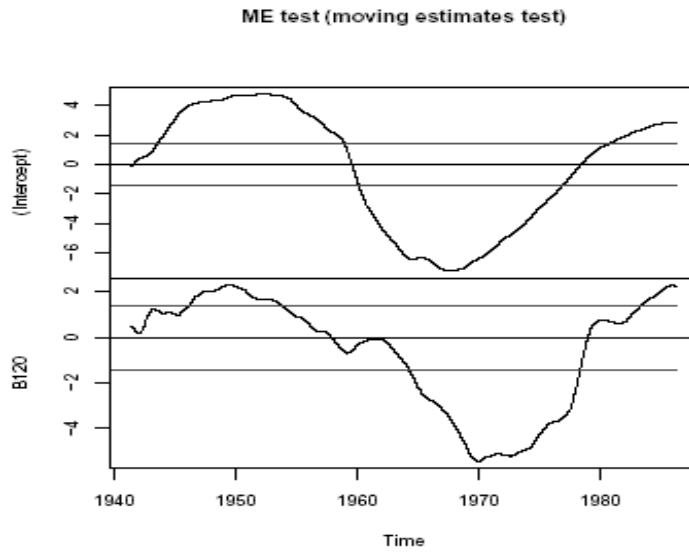


FIGURE 32—120-Month Value Weighted Return Horizon

**TABLE 3. OLS Quadratic Spectral Kernel (OLS-QS), Design Matrix Bootstrap (DMB) and Moving Blocks Bootstrap (MBB) of the First Order Autoregression Slope  $B_k$ ; Equally and Value Weighted Real Returns.**

	Return Horizon (K months)							
	12	24	36	48	60	72	96	120
Panel A: Equally-Weighted								
OLS-QS:								
$B_k$	-0.076 (0.189)	-0.253 (0.000)	-0.318 (0.020)	-0.354 (0.006)	-0.362 (0.004)	-0.169 (0.051)	0.010 (0.936)	0.055 (0.758)
DMB								
$B_k$	-0.137 (0.007)	-0.254 (0.000)	-0.317 (0.000)	-0.353 (0.000)	-0.362 (0.000)	-0.168 (0.000)	0.0105 (0.010)	0.053 (0.061)
MBB								
$B_k$	-0.143 (0.178) 10	-0.263 (0.113) 8	-0.322 (0.030) 9	-0.359 (0.017) 11	-0.368 (0.002) 10	-0.180 (0.084) 10	0.009 (0.888) 9	0.057 (0.451) 9
Optimal Number of Blocs								
Panel B: Value -Weighted								
OLS-QS:								
$B_k$	-0.0856 (0.288)	-0.255 (0.000)	-0.279 (0.165)	-0.155 (0.546)	-0.102 (0.696)	0.031 (0.823)	-0.015 (0.886)	-0.172 (0.000)
DMB								
$B_k$	-0.085 (0.224)	-0.254 (0.000)	-0.275 (0.000)	-0.152 (0.000)	-0.100 (0.009)	0.031 (0.267)	-0.015 (0.608)	-0.173 (0.000)
MBB								
$B_k$	-0.082 (0.424) 10	-0.251 (0.012) 9	-0.269 (0.016) 9	-0.147 (0.246) 10	-0.100 (0.412) 10	0.025 (0.716) 9	-0.022 (0.863) 9	-0.178 (0.041) 8
Optimal Number of Blocs								

**Note:** The tests are based on the following specification  $R_{i,t+k} = \alpha + B_k R_{i,t-k} + \epsilon_t$ . Figures in parentheses are  $p$ -values.

*B. Robust Test of Random Walk*

The autoregression in equation 1 was estimated for each return horizon by both *OLS* and *MBB* for the whole sample period and subperiods determined by Andrews (1993) and Andrews and Ploberger (1994) stability tests. The *OLS* standard errors are corrected for autocorrelation using *HH* estimator as in Fama and French (1988). Also, the test labeled *OLS-QS*, reports *OLS* quadratic spectral kernel standard errors suggested by Andrews (1991). These two tests sequentially provide *AC* and *HAC* standard errors but do not correct for sample bias. To do so, the *MBB* test described in section III B is estimated by setting an arbitrary numbers of blocks from 1 to 15.<sup>14</sup> The optimal number of blocks for each return horizon is selected by choosing the block size that minimizes the *RMSE* of the test variance based on 1000 resamples for each sample. Consistent with Hall, Horowitz, and Jing (1995) the optimal number of blocks sets around 10, which is the inverse of the number of observations cube root. To gain better interpretation about heteroskedasticity and autocorrelation effects, the paper also reports the result of *DMB (MBB1)*.<sup>15</sup> This test solves for small sample bias and provides *HC* standard error that dominant MacKinnon and White (1985) *HC* estimator.

Table 3 reports the results of random walk test for the period 1926 through 2001 using the tests described above. *MBB* (10) equally and value weighted return reports slightly lower slope and higher p-value than *OLS-QS*. However it is clear that there is no reliable evidence of predictability when *MBB* is used except for holding periods 2- and 3-year value weighted returns and 3- and 4-year equally weighted returns. For value weighted returns, the *p*-values at 2- and 3-year holding periods are 0.012 and 0.016 respectively, imply the rejection of the null hypothesis of random walk. The *p*-values for 3- and 4-year equally weighted returns are slightly lower than 0.05. For the rest of holdings period the hypothesis of random walk is accepted. The *p*-values are high for the null to be rejected.

Some practical implication concerning autocorrelation, heteroskedasticity and finite sample bias can be gleaned from our results. First, adjusting for heteroskedasticity and point estimate bias

---

14. When the number of blocks is set to one, i.e., the block size is 100% of the effective sample; the *MBB* becomes identical to *DMB* since *DMB* is a special case of *MBB*.

15. Hall, Horowitz, and Jing (1995) provide evidence that *DMB* dominates any version of White *HC* estimator.

**TABLE 4. Moving Blocks Bootstrap (MBB) of the First Order Autoregression Slope  $\beta_{K,Dj}$  ; CRSP Equally and Value Weighted Real Returns**

	Return Horizon ( $K$ months)			
	12	24	36	48
Panel A: Equally-Weighted				
<i>MBB:</i>				
$B_{K,1}$	-0.068 (0.502)	-0.259 (0.269)	-0.301 (0.152)	-0.352 (0.019)
$B_{K,2}$	-0.332 (0.072)	-0.223 (0.185)	-0.376 (0.010)	-0.469 (0.014)
$B_{K,3}$		-0.435 (0.001)	-0.501 (0.020)	
Sum	-0.399	-0.949	-1.179	-0.816
Wald Static	2.805 (0.094)	5.036 (0.024)	6.6429 (0.010)	7.227 (0.007)
Optimal Number of Blocs	10	10	10	10
Panel B: Value-Weighted				
<i>MBB:</i>				
$B_{K,1}$	-0.065 (0.719)	-0.371 (0.059)	-0.345 (0.009)	-0.201 (0.165)
$B_{K,2}$	-0.159 (0.132)	-0.272 (0.028)	-0.331 (0.058)	-0.280 (0.087)
$B_{K,3}$	0.201 (0.193)	-0.005 (0.962)	0.080 (0.540)	0.162 (0.307)
Sum	-0.023	-0.649	-0.595	-0.315
Wald Static	0.006 (0.934)	6.902 (0.009)	3.441 (0.064)	0.633 (0.426)
Optimal Number of Blocs	10	10	10	10
	Return Horizon ( $K$ months)			
	60	72	96	120
Panel A: Equally-Weighted				
<i>MBB:</i>				
$B_{K,1}$	-0.066 (0.504)	-0.173 (0.098)	0.005 (0.933)	0.032 (0.664)
$B_{K,2}$	-0.081 (0.199)	-0.273 (0.034)	-0.127 (0.152)	-0.112 (0.298)
$B_{K,3}$	-0.080 (0.786)			
Sum	-0.227	-0.447	-0.119	-0.079
Wald Static	0.459 (0.529)	4.238 (0.040)	0.810 (0.368)	0.203 (0.653)
Optimal Number of Blocs	9	10	9	9

(Continued)

TABLE 4. (Continued)

	Return Horizon ( $K$ months)			
	60	72	96	120
Panel B: Value-Weighted				
<i>MBB</i> :				
$B_{K,1}$	-0.096 (0.352)	-0.057 (0.588)	-0.036 (0.791)	0.032 (0.804)
$B_{K,2}$	-0.120 (0.050)	0.031 (0.935)	-0.125 (0.205)	-0.232 (0.013)
$B_{K,3}$	0.121 (0.696)	0.305 (0.017)	0.238 (0.012)	0.019 (0.878)
Sum	-0.096	0.279	0.076	-0.180
<i>Wald</i> Static	0.112 (0.737)	0.737 (0.391)	0.078 (0.780)	0.481 (0.488)
Optimal Number of Blocs	10	9	9	9

**Note:** The test is based on the following specification:

$$R_{t,t+K} = \alpha_K + \sum_i^n C_i \alpha_{K,i} + \sum_{j=1}^m \beta_{K,D_j} D_j R_{t-k,t} + \varepsilon_t. \text{ Figures in parentheses are } p\text{-values.}$$

but not autocorrelation in testing for the random walk in stock prices is strongly bias toward rejection the null. The resulting  $p$ -values from *DMB* estimation are too much small in comparison to that of *QS* and *MBB*. Second, consistent with *F&F*, bias adjustment is not large enough to disturb the results. The resulting *MBB*  $p$ -values are comparable to that of *QS* for the majority of holding periods. Finally, autocorrelation plays the most important roll in the estimation. The results can be gleaned from comparing *MBB* and *DMB*.

Table 4 reports the results of the *MBB* test for the autoregression specification in 10 after adjusting for structural break points reported in table 3. For equally weighted returns, the paper doesn't find any evidence of mean aversion behavior in stock prices, all the autocorrelations are negative across subperiods. As expected by the model, the stationary component of stock prices is slower than that estimated using the specification of *F&F*. The  $p$ -values of the *Wald* statistics for the hypothesis that the sum of the slopes are equal to zero for 2-, 3-, 4- and 6-year horizons are 0.249, 0.010, 0.007, and 0.004, respectively. This suggests the rejection of the random walk hypothesis for those horizons.

For value weighted portfolios, the paper reports positive autocorrelations only in the third subperiod (after 70's in general). This

suggests that the parameter of the structural break in the transitory component is negative for this subperiod. In contrast, the model expects that the absolute value for the sum of the autocorrelations will decline for those horizons that display mean aversion. (i.e., we expect weaker mean reversion behavior for value weighted returns).

Table 4 panel (B) reports the p-values of the *Wald* statistics for value weighted returns, the 3- and 10-year autocorrelations for value weighted returns is not significantly different from zero. However it is clear that there is no reliable evidence of predictability for value weighted portfolios except for the holding period of 2-year.

## VI. Conclusion

The results of this paper suggest that observed mean reversion in prior tests, based on simulated *OLS* first order autoregression, is understated due to a) size distortions resulting from the use of long-horizon returns; b) the non Gaussian distribution of returns; c) the clustering heteroskedasticity and autocorrelation in the overlapping returns; and d) the structural breaks in the data generating process. Three main results are revealed from the analysis of structural breaks. First, the most important structural break dates for those series' occurred on the subhorizon where the depression period returns are not overlapped. The strongest deviation for the considered tests indicate that the main shifts in entire parameter vector and individual parameters occur around the beginning of the 70's or around the mid 80's. Second, there is strong evidence that the overlapping returns display a second structural break where the depression period returns are overlapped. The *MOSUM* test indicates that the shift occurred around the beginning of 50's and the *ME* test clarifies that shift in the 50's is generally related to the intercept while the shift in the slope of the autoregression occurs at the beginning of the effective sample (the mid 40's). Third, the structural breaks in the intercept are horizon-independent while the shift in the slopes are monotonic with horizons. After adjusting for structural break in the long-horizon returns, the *MBB* test rejects the random walk hypothesis across 2-, 3-, 4-, and 6-year horizons for equally weighted portfolios. However, it is clear that there is no reliable evidence of predictability for value weighted portfolios except for the holding period of 2-year. Adaptive point estimates of the autocorrelation parameters all give strong evidence of heavy size distortions and support the use of robust estimation method.

**References**

- Andrews, D. W. K. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59: 817–858.
- Andrews, D. W. K. 1993. Tests for parameter instability and structural change with unknown change points. *Econometrica* 61: 821–856.
- Andrews, D. W. K., and Ploberger, W. 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62: 1383–1414.
- Chu, C.- S. J.; Hornik, K.; and Kuan, C.- M. 1995a. MOSUM tests for parameter constancy. *Biometrika* 82: 669–720.
- Chu, C.- S. J.; Hornik, K.; and Kuan, C.- M. 1995b. The moving-estimates test for parameter stability. *Econometric Theory* 11: 669–720.
- Chu, C.- S. J.; Stinchcombe, M.; and White, H. 1996. Monitoring structural change. *Econometrica* 64: 1045–1065.
- Cochrane, J. 1988. How big is the random walk in GNP? *Journal of Political Economy* 96: 893–920.
- Efron, B. 1979. Bootstrap methods. Another look at the jackknife. *Annals of Statistics* 7: 1–26.
- Efron, B. 1982. The jackknife, the bootstrap and other resampling plans. SIAM Philadelphia.
- Efron, B. and Tibshirani, R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1: 54–77.
- Fama, E. F., and French, K. R. 1988. Permanent and temporary components of stock prices. *Journal of Political Economy* 96: 246–273.
- Fitzberger, B. 1998. The moving blocks bootstrap and robust inference for linear least squares and quintile regressions. *Journal of Econometrics* 82: 235–287.
- Hall, P.; Horowitz, J. L.; and Jing, B. Y. 1995. On blocking rules for the bootstrap with dependent data. *Biometrika* 82: 561–574.
- Hansen, L. P., and Hodrick, R. J. 1980. Forwards exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy* 88: 829–853.
- Horowitz, J. L., and Savin N. E. 1992. Noninvariance of the Wald test: the bootstrap to the rescue, Working Paper no. 92–04: University of Iowa.
- Kendall, M. 1954. Note on bias in the estimation of autocorrelation. *Biometrika* 41: 403–404.
- Kim, M.; Nelson, C.; and Startz, R. 1991. Mean reversion in stock prices? A reappraisal of the empirical evidence. *Review of Economic Studies* 58: 515–528.
- Kusch, H. R. 1989. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17: 1217–1241.
- Leisch, K.; Hornik, M.; and Kuan, C.- M. 2000. Monitoring structural changes with the generalized fluctuation test. *Econometric Theory* 16: 835–854.
- Lo, A., and MacKinlay, A. 1988. Stock market prices do not follow random

- walks: Evidence from a simple specification test. *The Review of Financial Studies* 1: 897–916.
- MacKinnon, J. G., and White, H. 1985. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* 29: 305–326.
- Liu, R. Y. 1988. Bootstrap procedures under some non-i.i.d. models. *Annals of Statistics* 16: 1696–1708.
- Liu, F., and Singh, K. 1992. Moving blocks jackknife and bootstrap capture weak dependence. In D. LePage, R. and L. Billard (eds). *Exploring the Limits of Bootstrap*. New York: John Wiley.
- Mammen, E. 1993. Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics* 21: 255–285.
- McQueen, G., 1992. Long-horizon mean-reverting stock prices revisited. *Journal of Financial Economics* 27: 1–18.
- Noreen, E. W. 1989. *Computer-Intensive Methods for Testing Hypotheses: An Introduction* New York: John Wiley & Sons.
- Newey, W. K., and West, K. D. 1987. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55: 703–708.
- Perron, P. 1989. The great crash, the oil-price shock, and the unit-root hypothesis. *Econometrica* 57: 1361–1402.
- Ploberger, W., and Kramer, W. 1992. The CUSUM test with OLS residuals. *Econometrica* 60: 271–285.
- Ploberger, W.; Kramer, W.; and Kontrus, K. 1989. A new test for structural stability in the linear regression model. *Journal of Econometrics* 40: 307–318.
- Poterba, J., and Summers, L. 1988. Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics* 22: 27–59.
- Richardson, M., and Stock, J. H. 1989. Drawing inferences from statistics based on multiyear asset returns. *Journal of Financial Economics* 25: 323–348.
- Richardson, M. 1993. Temporary components of stock prices: A skeptic's view. *Journal of Business and Economic Statistics* 11: 199–207.
- Singh, K. 1981. On the asymptotic accuracy of Efron's bootstrap. *Annals of Statistics* 9: 1187–1195.
- Summers, L. 1986. Does the stock market rationally reflect fundamentals values? *The Journal of Finance* 41: 591–601.
- Turner, C. M.; Startz, R.; and Nelson, C. R. 1989. A Markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics* 25: 3–22.
- Vicereira M. 1997. Testing for Structural Change in the Predictability of Asset Returns. Manuscript: Harvard University.
- White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48: 817–838.
- Wu, C. F. J. 1986. Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics* 14: 1261–1295.